Title: Envelope responses in single-trial EEG indicate attended speaker in a "cocktail party"

Abbreviated Title: Envelope responses indicate attended speaker

Authors: Cort Horton[1], Ramesh Srinivasan[1,2], and Michael D'Zmura[1]

[1]Department of Cognitive Sciences and [2]Department of Biomedical Engineering, University of California, Irvine, CA 92697

Corresponding Author: Cort Horton, Department of Cognitive Sciences, 2201 Social & Behavioral Sciences Gateway Building, University of California, Irvine, CA 92697-5100. Email: chorton@uci.edu

Grants: This work was supported by grants from the National Institutes of Health (2R01-MH68004) and the Army Research Office (ARO 54228-LS-MUR).

Note: This is a working paper dated 6/20/13, and is not yet peer-reviewed or published

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **20 JUN 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE **Envelope responses in single-trial EEG indicate attended speaker in a ’cocktail party’**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Irvine,Department of Cognitive Sciences,Irvine,CA,92697**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Recent studies have shown that neural activity in auditory cortex encodes the envelope of speech, and that this encoding is more robust for attended speech than for unattended speech during multi-speaker (?cocktail party?) situations. To determine if this effect could form the basis for a novel brain-computer interface (BCI), we investigated the accuracy with which a subject?s locus of attention during a ?cocktail party? task could be ascertained from envelope responses present within single trials of EEG. We found that the attended speaker can be determined reliably from short periods of EEG, with accuracy improving as a function of trial length. Furthermore, we compared the performance of this envelope-based attention classifier to others based on changes in steady-state responses (elicited via 40 and 41 Hz amplitude modulations of the speech) and hemispheric lateralization of alpha power. We found that the neural responses to the speech envelopes were far more robust indicators of attention than the others. These results suggest that envelope-related signals recorded in EEG data can be used to form robust auditory BCI?s that do not require artificial manipulation (e.g., amplitude modulation) of stimuli in order to function.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**21**	

Abstract

Recent studies have shown that neural activity in auditory cortex encodes the envelope of speech, and that this encoding is more robust for attended speech than for unattended speech during multi-speaker ("cocktail party") situations. To determine if this effect could form the basis for a novel brain-computer interface (BCI), we investigated the accuracy with which a subject's locus of attention during a "cocktail party" task could be ascertained from envelope responses present within single trials of EEG. We found that the attended speaker can be determined reliably from short periods of EEG, with accuracy improving as a function of trial length. Furthermore, we compared the performance of this envelope-based attention classifier to others based on changes in steady-state responses (elicited via 40 and 41 Hz amplitude modulations of the speech) and hemispheric lateralization of alpha power. We found that the neural responses to the speech envelopes were far more robust indicators of attention than the others. These results suggest that envelope-related signals recorded in EEG data can be used to form robust auditory BCI's that do not require artificial manipulation (e.g., amplitude modulation) of stimuli in order to function.

Keywords

Selective attention, speech envelopes, brain-computer interfaces, alpha lateralization, steady-state responses

Introduction

A great deal of effort has been devoted to mapping out the relationship between the acoustic properties of speech utterances and their associated neural responses. At the level of the cortex, the feature in speech that seems to be best represented is its temporal envelope. Researchers have found strong correlations between auditory cortical activity and speech envelopes [1–3] which appear to be driven by endogenous oscillations shifting in phase to track the slow (< 10 Hz) amplitude modulations present in the envelope [4,5]. This phase-locking was originally thought to reflect a feed-forward process, but recent studies have found that phase-locking is also subject to top-down factors. For example, phase-locking is diminished when speech is unintelligible [6,7], and is strengthened when the speaker's face is visible [8]. Additionally, in situations with multiple competing talkers, such as the classic "cocktail party" task [9], the auditory system preferentially phase-locks to the envelope of the attended speech [10], and attempts to remain out of phase with the envelope of competing speech [11].

If this difference in phase-locking between attended and unattended speech is visible in single-trial electroencephalographic (EEG) data, it may be possible to develop a novel brain-computer interface (BCI) that acts upon neural responses to the envelopes of multiple simultaneously-presented auditory stimuli. BCIs vary widely in their implementation, but the common goal is to use brain-generated signals to communicate or control a computer interface [12]. Some BCIs require users to modulate their brain activity, such as motor rhythms, in order to signal intent [13], but these often require considerable training. Other BCIs forgo training and instead have subjects make choices by attending to one of multiple visual and/or auditory stimuli. By presenting each stimulus at a different time [14] or frequency [15,16], evoked responses to each can be extracted and compared for signs of attention. While this method allows for complex interfaces, such as a full BCI-controlled keyboard [14], it is constrained by the need for artificial stimuli (*e.g.*, flickered or modulated). An envelope-based BCI could operate on more naturalistic auditory stimuli, such as speech or music. For example, an envelope-based BCI could interface with a hearing aid in the hopes of adjusting the relative volumes of competing speakers during "cocktail party" scenarios based on which speaker the user is attending. These scenarios are common in everyday life, and are particularly problematic for people with hearing impairment [17]. Thus, such a BCI could potentially provide enormous benefit.

We determined in the present study whether the differences in phase-locking to the envelopes of attended and unattended speech could be reliably observed in single-trial EEG data and so provide the basis for an envelope-based BCI. Adult subjects engaged in a cocktail party task in which they attended to one of two competing speakers while high-density EEG data were recorded. We used the EEG data to extract phase-locked responses to the envelopes of both speakers, and then assessed our ability to decode their side of attention as a function the duration of EEG data used to extract the responses. We found that envelope responses were sufficiently represented in EEG to decode side of attention from brief segments of data, with accuracy improving as data segment duration increased.

Furthermore, we wanted to compare classification of attention lateralization using envelope-locked responses to classification of attention lateralization using other indicators that have appeared in BCIs. First, since the speakers were located in different places, we expected to see signs of attention in the EEG data's spectral content. The deployment of attention to the left or right side of space is associated with hemispheric lateralization of oscillatory power in several frequency bands [18–20]--particularly in the alpha band (8-12 Hz). Alpha power lateralization can be sufficiently robust to discriminate a subject's side of attention without further need to consider any stimulus-related brain activity [20,21]. Second, some BCIs decode attention from changes in auditory steady-state responses (ASSRs) [16], as attention has been shown to boost ASSR magnitudes [22]. Thus, we amplitude modulated the left and right speech streams at 40 and 41 Hz in order to induce ASSRs in the EEG data. We found that classification accuracy using envelope responses greatly outperformed classification using either alpha lateralization or ASSR magnitude, further reinforcing the potential for envelope-based BCIs.

Methods

The data were also used in a previous study examining how cortical entrainment to speech envelopes is involved in selectively attending to one of multiple speakers [11]. The current study shared some of its data pre-processing steps, but otherwise had distinct goals and analyses.

Participants

All experimental procedures were approved by the Institutional Review Board of the University of California, Irvine. Ten young adults (2 female) between the ages of 21 and 29 volunteered to participate in the study, although one had to be excluded due to excessive EEG artifacts. All reported having normal hearing and no history of neurological disorder. Written informed consent was obtained from each subject prior to participation in the study.

Task and Stimuli

Each participant sat in a sound-attenuated testing chamber and faced a computer monitor that was flanked on either side by a loudspeaker (Fig 1). Before each trial, the subject was presented with a visual cue to attend to either the left or right speaker (chosen at random) while maintaining visual fixation on a cross in the center of the monitor. During the trial, the left and right speakers played independent speech stimuli consisting of a series of spoken sentences taken from the TIMIT speech corpus [23]. To build these speech stimuli, sentences were drawn from the corpus at random and concatenated until the total length of each channel exceeded 22 seconds, with silent gaps longer than 300 ms being reduced to 300 ms. No sentence was reused within experimental sessions. Envelopes for the stimuli were calculated by bandpass filtering the result of a Hilbert transform using a passband of 2 to 30 Hz. After constructing the stimuli, the left and right channels were sinusoidally amplitude-modulated at 40 and 41 Hz, respectively, in order to induce ASSRs. These modulation frequencies induce robust ASSRs [24,25]and do not interfere with the intelligibility of the speech [26–28]. At the end of each trial, subjects were shown the transcript of a sentence from the trial, and were required to indicate via a button press whether the sentence was played on the attended side. In practice, this task was very difficult unless subjects ignored the unattended side completely, as the memory load required to maintain both sides was prohibitive. Subjects were allowed to practice the task until their performance exceeded 80%, and were required to maintain that level throughout the experiment. Subjects completed 320 trials each (8 blocks, 40 trials per block), spread over 1 to 2 weeks, with the exception of one subject who only completed 240 trials due to equipment failure.

EEG Recording and Pre-Processing

During the task, we recorded 128 channels of EEG using caps, amplifiers, and software produced by Advanced Neuro Technology. Electrodes were placed following the international 10/5 system [29], and all channel impedences were kept below 10 kΩ. The EEG data was sampled at 1024 Hz with an online average reference but no online filters. After the experiment, EEG data were exported into MATLAB (MathWorks, Natick, MA) for all further processing and analyses.

Each channel of EEG was Butterworth filtered with a pass band of 1 to 50 Hz. Filters were run both forwards and in reverse to eliminate phase shifts. The filtered data were then down-sampled to 256 Hz and segmented into individual trials which were 20 seconds long, beginning one second after the onset of the sentences. The delay between sentence onset and analysis window onset was necessary because neural onset responses are known to be large relative to envelope-related activity [1,2]. Furthermore, since the left and right speech began simultaneously, they very briefly had correlated envelopes, which could impair later analyses. The segmented trials were visually inspected to exclude those with excessive artifacts (mean 16.6 trials per subject). The remaining data were then entered into the Infomax Independent Component Analysis algorithm available as part of the EEGLAB toolbox [30].

Components corresponding to artifacts such as eye movements and muscle activity were removed [31], and all remaining components were projected back into channel space for subsequent analyses.

EEG Feature Extraction

We extracted three different features from the EEG to use in the classification of attention. First, we obtained envelope-related responses by calculating the cross-correlation functions [32] between each stimulus's envelope and the EEG channels. Cross-correlation measures the similarity of two time-series as a function of lag between them, and has been used previously to quantify neural responses to speech envelopes [1,2,11]. For discrete functions f and g, it is defined as:

$$(f \star g)(n) = \sum_{m=-\infty}^{\infty} \frac{f[m]g[n+m]}{\sigma_f \sigma_g},$$

in which σ_f and σ_g are the standard deviations of f and g. The second feature we extracted from each channel was a measure of power in the alpha band, calculated by Fourier transforming each trial and then summing power across all of the frequency bins between 8 and 12 Hz. The final feature we extracted from each channel was a measure of the ASSRs, calculated by again Fourier transforming each trial, but taking just the frequency bins corresponding to the left (40 Hz) and right (41 Hz) modulators.

Classification

Classification was run independently on each subject's data, and on each of the three EEG signals of interest, using the linear discriminant classifier built into MATLAB. The classifier attempts to use a set of training data to build an optimal hyperplane that separates the two conditions. Once built, a different set of data can be used to test how well that model accounts for new data. To form the training and testing sets, we divided each subject's trials randomly, using 75% of the trials to train the classifier, and 25% of trials to test it. The accuracy of those predictions comprised a single estimate of the classifier's performance. We then repeated the process using a new random split of training and test trials, until we had 500 such estimates. We used the mean of those estimates to gauge the overall performance of the classifier. Classification accuracy for a given trial length was stated to be significantly above chance (one-way, α=.05) if the 5^{th} percentile of those accuracy estimates exceeded 50%. We then implemented a Bonferroni correction to account for multiple comparisons at the six different trial lengths, which changed the threshold for significance for a single trial length to be set at the $5/6^{th}$ percentile of the distribution. A graphical representation of the classification process appears in Fig 2.

For classification algorithms, it is necessary to balance the number of features/variables submitted to the classifier with the number of trials in the training set. Thus, we only submitted the most informative 15 channels to the classifier for each EEG signal of interest. For alpha power, we found those channels by subtracting the mean alpha power of "attend left" trials in the training set from the mean alpha power of "attend right" trials in the training set, and then selected those channels where the differences were greatest (Fig 3 left). For the ASSRs, we used the 15 channels where the response magnitudes were greatest in the training data, averaged across the two modulation frequencies (Fig 3 right). For the envelope cross-correlations, we first found the 15 channels where the differences were largest between the mean attended and mean unattended cross-correlation functions in the training data. We then found the latencies of the peaks in that difference function, indicating points where the cross-correlations of attended and unattended speech envelopes should be most distinct from one another (Fig 4). Thus, for each trial the classifier was given the cross-correlation value of those 15 channels at each of these peak latencies.

To evaluate the amount of data needed for successful classification, we varied the length of length of the trials fed into the classifier. For trial lengths shorter 20 seconds (the length of trials in the original dataset), we divided each trial into multiple shorter trials (i.e. one 20 seconds trial cut into five 4 second trials). Each of the EEG measures (envelope-cross correlations, alpha power, and ASSRs) was then calculated on these new shorter trials. For trial lengths longer than 20 seconds, we concatenated the EEG of multiple trials from the same condition.

Results

Behavior

Participants were able to exceed the required performance on the behavioral task throughout all experimental sessions (mean 82.45% correct). They reported the task as being challenging due to the effort required to maintain all of the novel sentences from the attended side in working memory.

EEG Feature Extraction

Using the training data for each subject, we calculated envelope cross-correlation functions that mirrored those observed in other studies [1–3,11]. The channels where the attended and unattended cross-correlation functions were most distinct were located over frontal and temporal sites, which we have previously identified as consistent with sources in both early and later auditory areas [11]. Most subjects showed three distinct peaks in the difference function between the attended and unattended cross-correlations, as depicted in figure 4. The latencies of those peaks corresponded to the latencies of well-known auditory evoked responses [33].

Alpha power in each subject's training data showed the expected pattern of hemispheric lateralization differences between "attend left" and "attend right", although those differences were generally weak. The channels selected for classification in each subject were primarily located over parietal cortex, but also included some neighboring occipital and temporal electrodes. A topographic plot of a representative subject's alpha power differences appears in figure 3 (left).

Robust ASSRs were present in each subject's training data, but these ASSRs did not show signs of being modulated by attention in any subject. The largest responses were recorded over frontal, occipital, and posterior temporal sites – consistent with previous studies [24,34]. A topographic plot of a representative subject's average ASSRs appears in figure 3 (right).

Classification

We found that cross-correlation functions calculated from single trials of EEG were highly effective in decoding which speaker had been attended, with classification accuracy exceeding chance for all subjects at all tested trial lengths (Fig 5). At the shortest trial length tested, 2 seconds, the average classification performance across subjects was 63%. That performance increased monotonically as trial length increased, reaching 75% accuracy on average across subjects with 10 seconds of data. At 40 seconds, we saw evidence of ceiling effects on classifier performance, with one subject at 100% classification accuracy. Classification accuracy differed greatly across subjects, but those differences remained consistent across all trial lengths (i.e. the best subjects at 2 second trials were also the best at other trial durations).

In contrast to the classification using cross-correlations, we found that classification based on alpha power lateralization was poor. Classification performance never dipped below 50% for any subject or trial length, indicating that there was some information about attention available in the alpha power. However, only a few subjects showed significantly above chance classification accuracy, and those were exclusively found at short trial lengths. Classification based on ASSR magnitudes performed even worse, with accuracy never exceeding chance for any subject at any trial length.

Discussion

Cross-Correlations

We were able to determine subjects' locus of attention using the cross-correlations between the speech envelopes and their EEG, with accuracy increasing as a function of trial length. Encouragingly, classification performance for short trial lengths was on par with or exceeded that seen in a comparable recent study using magnetoencephalography (MEG) [35], a technology that is often used to measure speech responses but that would be far less practical for BCI purposes.

The accuracy with which we were able to classify attention varied widely across subjects, with as much as a 25% difference in accuracy between the best and worst subjects at longer trial lengths. Put another way, the classification accuracy using 2 seconds worth of EEG data from the best subject was equivalent in performance to 20 seconds of data from the worst subject. These differences between subjects are similar to those seen in other types of BCIs, where it has long been known that some subjects innately perform better with BCIs than others, and pre-training ability to use a BCI is a very strong predictor of post-training success [36]. However, in this task the individual differences would not be driven by a failure to learn how to modulate certain brain rhythms, but rather on differences in the robustness of the stimulus-related signals of interest. On *post-hoc* examination of the high and low performing subjects, the better performers had stronger cross-correlations between the speech envelopes and their EEG, and thus also would have had higher signal-to-noise ratios on individual trials. Although we did not observe any notable differences between these groups in their behavioral performance, it would be interesting in future work to find out if their abilities diverged during more challenging multi-talker tasks.

Alpha Lateralization

Since alpha power showed the expected lateralization in the subject averages, it may seem puzzling at first why classification based on alpha power in single trials was ineffective. The most likely explanation is that alpha lateralization is associated with the *deployment* of spatial attention, not the *maintenance* of spatial attention. In this task, the most crucial time for deployment of spatial attention is at the very beginning of the trial, which is not included in our analysis window due to the problems that onset responses cause for the cross-correlation analyses. During our analysis window, the subjects are primarily maintaining attention at the cued location, which may not produce strong lateralization in alpha power. In fact, a similar cocktail party study found that alpha lateralization peaked 400 to 600 ms after sentence onsets, and was largely gone by 1000 ms (when our analysis window began) [37]. Subjects may have needed to briefly redeploy spatial attention at the transitions between sentences, which could explain why classification was able to exceed chance for a few subjects at short trial lengths and why the alpha power was lateralized in the subject averages. However, this lateralization was clearly not robust enough in single trials to produce useful classification of attention.

Additionally, it is important to note that spatial location was not the only cue available for distinguishing between the two competing speech streams. Once the target speech stream had been segregated from the competitor, there are many other features besides spatial location that subjects can use to track the target speech stream, including pitch, timbre, and tempo. If the speaker's voice had been the same on both the left and the right, the spatial feature would likely have been much more salient to the subjects, and consequently may have produced much stronger lateralization of alpha power.

Auditory Steady-State Responses

Attention did not affect the magnitudes of the ASSRs in the subject averages, and so it was unsurprising that the ASSR magnitudes did not help to classify attention in single trials. ASSR insensitivity to attention was also reported in a recent similar study [10]. Since ASSRs have been shown to be sensitive to attention in the past [22,38], and have been used to control a brain-computer interface [39], why were they not sensitive to attention here? We believe the difference lies in the fact that our stimuli were modulated speech utterances, whereas the studies in which ASSRs are affected by attention have used modulated tones, noise, or click trains. Modulated speech elicits much smaller ASSRs than modulated tones, noise, or reversed speech [25], suggesting that processing meaningful speech requires a suppression of the uninformative (and possibly interfering) amplitude modulation.

Conclusion

We have shown that neural responses to the envelopes of natural speech can be used to determine subjects' locus of attention, and thus could form the basis of a novel BCI. While the classification performance that we observed indicates that this BCI would not improve upon the information transfer rate in others, it would have the advantages of not requiring any training on the part of the subjects, and could be used with complex naturalistic stimuli such as speech. Additionally, while we only tested two-way classification, we could potentially increase the information transfer rate by increasing the number of speakers in the environment. In true cocktail-party scenarios, there may be dozens of competing speakers in the room, yet people are able to isolate the speaker of their choice. If people can do this behaviorally, there is good reason to believe that we could similarly isolate the neural response to the attended speaker, too. Thus, the upper limit of the information transfer rate for this BCI may be determined by the number of uncorrelated speech stimuli that can be played at once in a BCI system.

References

[1] Aiken S J and Picton T W 2008 Human cortical responses to the speech envelope. *Ear and Hearing* **29** 139–57

[2] Abrams D A, Nicol T, Zecker S and Kraus N 2008 Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *The Journal of Neuroscience* **28** 3958–65

[3] Lalor E C and Foxe J J 2010 Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *The European Journal of Neuroscience* **31** 189–93

[4] Luo H and Poeppel D 2007 Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54** 1001–10

[5] Giraud A-L and Poeppel D 2012 Cortical oscillations and speech processing: emerging computational principles and operations *Nature Neuroscience* **15** 511–7

[6] Ahissar E, Nagarajan S S, Ahissar M, Protopapas A, Mahncke H and Merzenich M M 2001 Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America* **98** 13367–72

[7] Peelle J E, Gross J and Davis M H 2012 Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral cortex (New York, N.Y. : 1991)*

[8] Zion Golumbic E, Cogan G B, Schroeder C E and Poeppel D 2013 Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33** 1417–26

[9] Cherry E C 1953 Some experiments on the recognition of speech, with one and with two ears *Journal of the Acoustical Society of America* **25** 975–9

[10] Ding N and Simon J Z 2012 Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology* **107** 78–89

[11] Horton C, D’Zmura M and Srinivasan R 2013 Suppression of Competing Speech through Entrainment of Cortical Oscillations. *Journal of neurophysiology*

[12] Birbaumer N and Cohen L G 2007 Brain-computer interfaces: communication and restoration of movement in paralysis. *The Journal of Physiology* **579** 621–36

[13] Wolpaw J R and Mcfarland D J 1994 Multichannel EEG-based brain-computer communication *Electroencephalography and Clinical Neurophysiology* **90** 444–9

[14] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology* **70** 510–23

[15] Kim D-W, Cho J-H, Hwang H-J, Lim J-H and Im C-H 2011 *A vision-free brain-computer interface (BCI) paradigm based on auditory selective attention* vol 2011 (IEEE)

[16] Desain P, Hupse A M G, Kallenberg M G J, Kruif B J De and Schaefer R S 2006 Brain-Computer Interfacing using Selective Attention and Frequency-Tagged Stimuli *Proceedings of the 3rd International BrainComputer Interface Workshop Training Course* 2–3

[17] Bronkhorst A W 2000 The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions *Acustica* **86** 117–28

[18] Worden M S, Foxe J J, Wang N and Simpson G V 2000 Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20** RC63

[19] Thut G, Nietzel A, Brandt S a and Pascual-Leone A 2006 Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26** 9494–502

[20] Thorpe S, D'Zmura M and Srinivasan R 2012 Lateralization of frequency-specific networks for covert spatial attention to auditory stimuli. *Brain Topography* **25** 39–54

[21] Srinivasan R, Thorpe S, Deng S, Lappas T and Zmura M D 2009 Decoding Attentional Orientation from EEG Spectra *Lecture Notes in Computer Science* pp 176–83

[22] Ross B, Picton T W, Herdman A T and Pantev C 2004 The effect of attention on the auditory steady-state response. *Neurology and Clinical Neurophysiology* **2004** 1–4

[23] Garofolo J, Lamel L, Fisher W, Fiscus J, Pallett D, Dahlgren N and Zue V 1993 TIMIT acoustic-phonetic continuous speech corpus

[24] Picton T W, John M S, Dimitrijevic A and Purcell D W 2003 Human auditory steady-state responses *International Journal of Audiology* **42** 177–219

[25] Deng S and Srinivasan R 2010 Semantic and acoustic analysis of speech by functional networks with distinct time scales. *Brain Research* **1346** 132–44

[26] Miller G A and Licklider J C R 1950 The Intelligibility of Interrupted Speech *The Journal of the Acoustical Society of America* **22** 167–73

[27] Drullman R, Festen J M and Plomp R 1994 Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America* **95** 2670–80

[28] Drullman R, Festen J M and Plomp R 1994 Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America* **95** 1053–64

[29] Oostenveld R and Praamstra P 2001 The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology* **112** 713–9

[30] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134** 9–21

[31] Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown M J, Iragui V and Sejnowski T J 2000 Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37** 163–78

[32] Bendat J S and Piersol A G 1986 *Random Data: Analysis and Measurement Procedures* vol 3 (John Wiley and Sons)

[33] Picton T W, Hillyard S A, Krausz H I and Galambos R 1974 Human auditory evoked potentials. I. Evaluation of components. *Electroencephalography and Clinical Neurophysiology* **36** 179–90

[34] Herdman A T, Lins O, Van Roon P, Stapells D R, Scherg M and Picton T W 2002 Intracerebral sources of human auditory steady-state responses. *Brain Topography* **15** 69–86

[35] Koskinen M, Viinikanoja J, Kurimo M, Klami A, Kaski S and Hari R 2012 Identifying fragments of natural speech from the listener's MEG signals. *Human brain mapping* **1489** 1477–89

[36] Neumann N and Birbaumer N 2003 Predictors of successful self control during brain-computer communication. *Journal of neurology, neurosurgery, and psychiatry* **74** 1117–21

[37] Kerlin J R, Shahin A J and Miller L M 2010 Attentional gain control of ongoing cortical speech representations in a "cocktail party". *The Journal of Neuroscience* **30** 620–8

[38] Müller N, Schlee W, Hartmann T, Lorenz I and Weisz N 2009 Top-down modulation of the auditory steady-state response in a task-switch paradigm. *Frontiers in Human Neuroscience* **3** 1–9

[39] Kim D-W, Hwang H-J, Lim J-H, Lee Y-H, Jung K-Y and Im C-H 2011 Classification of selective attention to auditory stimuli: toward vision-free brain-computer interfacing. *Journal of Neuroscience Methods* **197** 180–5

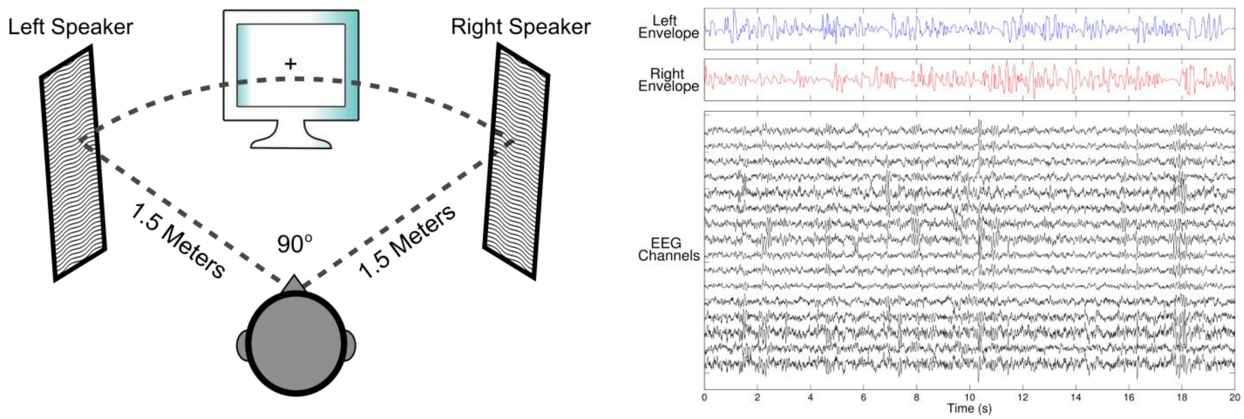

Figure 1: Experiment Design

Left: Layout of the equipment during the task (after [11]). Right: Example of raw data from one trial, showing the filtered envelopes of the left and right channel speech stimuli, as well as a subset of the EEG channels.

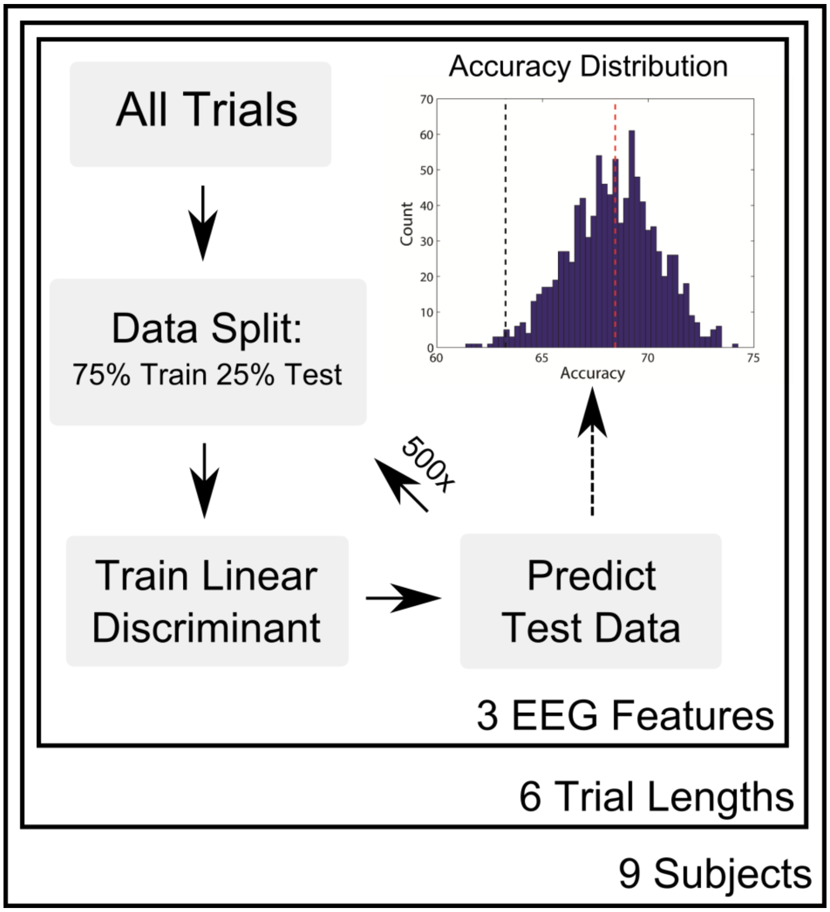

Figure 2: Classification Process

A graphical representation of the classification process. For a given subject, trial length, and EEG feature (cross-correlations, alpha lateralization, or ASSRs), trials are randomized into training and test sets. A linear discriminant is formed using the training data,which is then used to predict the attention condition of each test-set trial. The process is repeated 500 times, with new random splits of training and test trials. The accuracies of all iterations form a distribution (ex upper right). The mean of the distribution (red dashed line) is reported as the overall classification performance, and the accuracy is stated to be significantly above chance if the 5/6th percentile of the distribution (black dashed line) is above 50%.

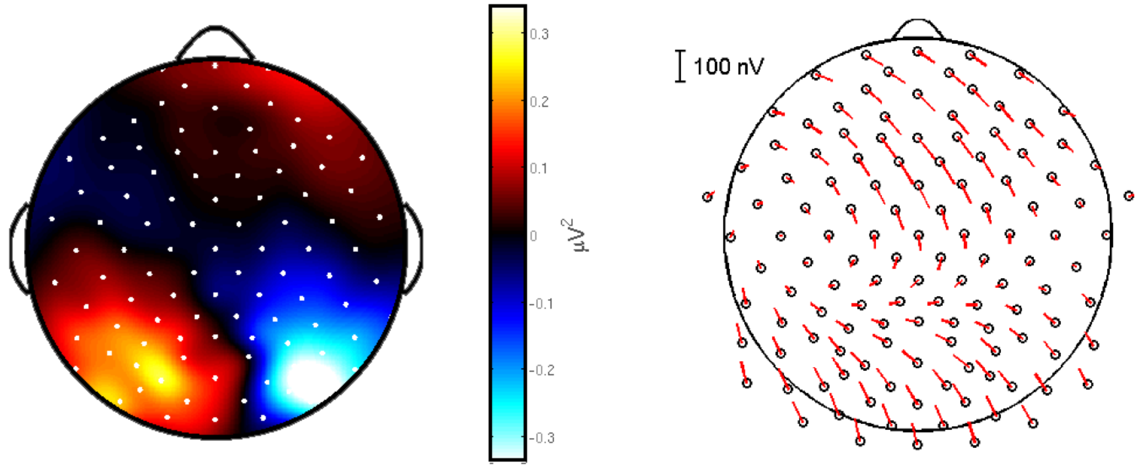

Figure 3: Alpha Lateralization and ASSRs

Left: A topographic plot for a representative subject showing the difference in alpha (8-12 Hz) power between the “Attend Left” condition and the “Attend Right” condition. The differences in power are maximal over parietal electrodes. Right: The average ASSRs for a representative subject. Magnitude is indicated by the length of the line extending from each electrode, while the phase is indicated by the angle. ASSR topography was typical for EEG studies, with peaks in magnitude over frontal and occipital electrodes.

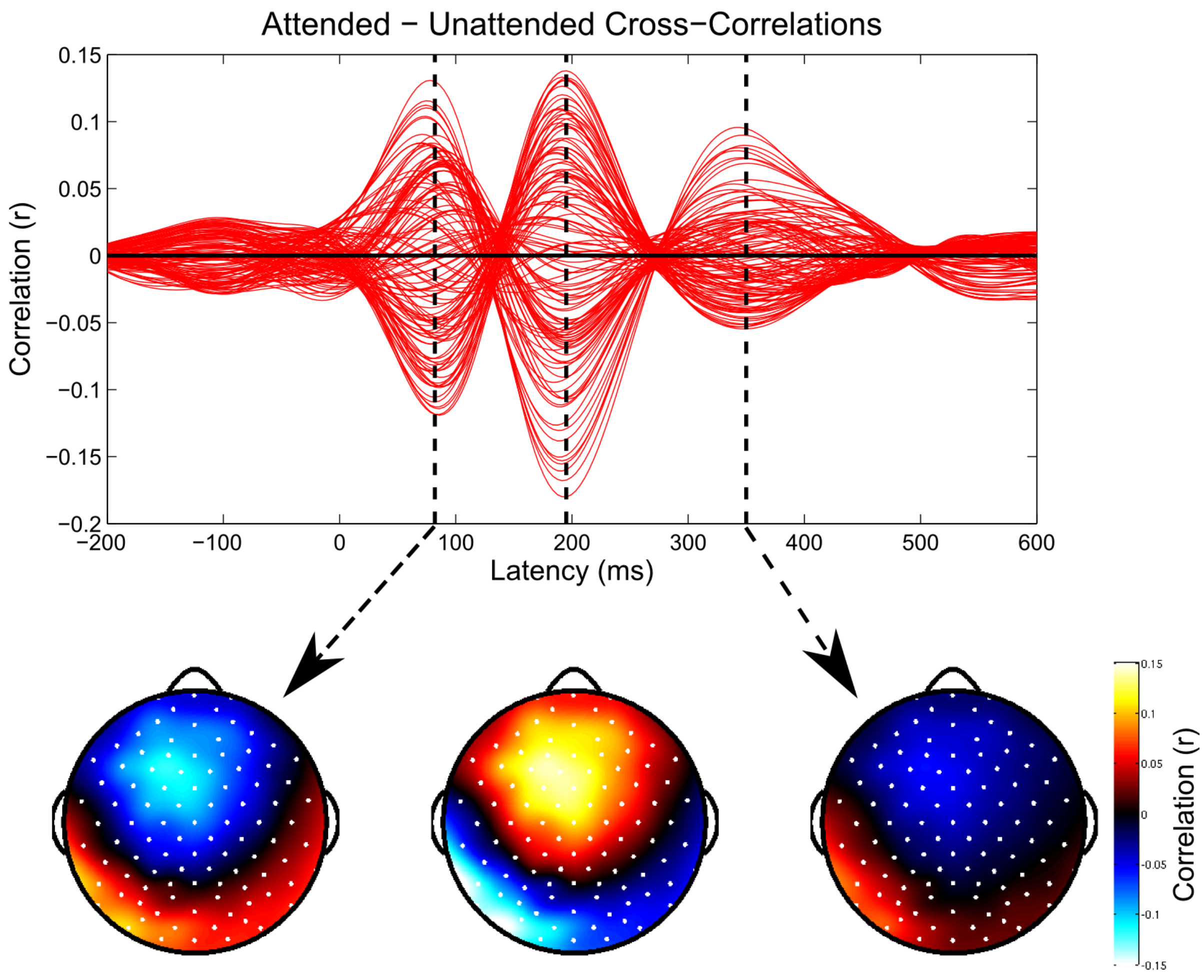

Figure 4: Cross-Correlations: Identifying Channels and Latencies of Interest

Top: For each subject, we calculate the average cross-correlation functions for the attended and unattended stimuli in the training set, and then plot their difference to identify latencies where they are most distinct. Bottom: At each latency of interest, we use the 15 channels with the largest magnitude differences between the attended and unattended cross-correlation functions. The position of those electrodes for one representative subject can be inferred from the scalp topographies of those differences. After [11].

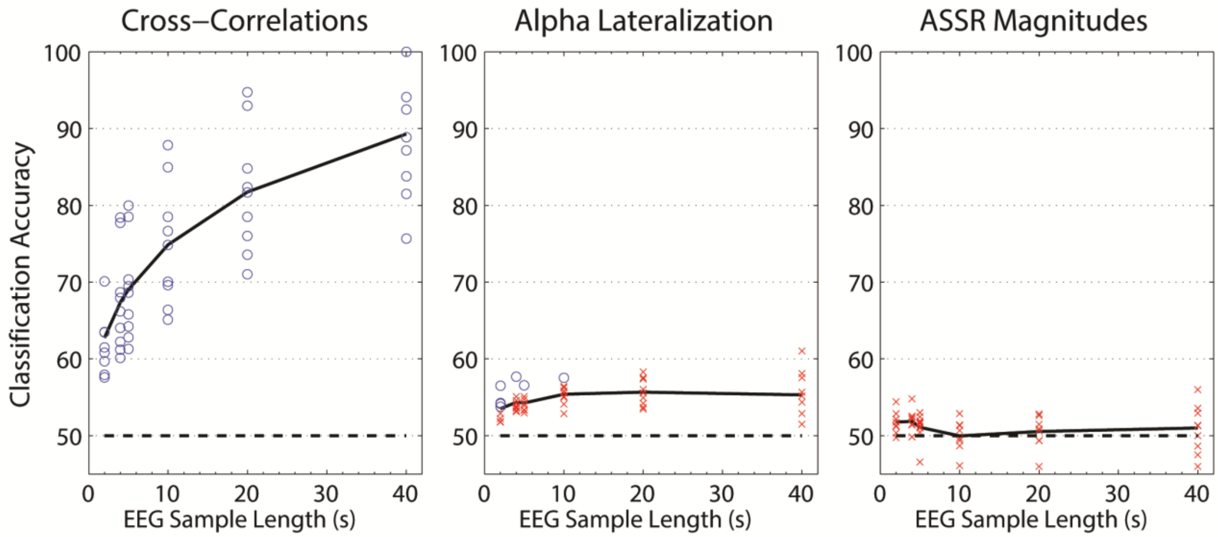

Figure 5: Classification Results

Classification accuracy is plotted as a function of EEG sample length for each of three different classifiers that make use of different features in the EEG. The mean accuracies for each subject are indicated by the individual data points, with the subject mean indicated by the black line. Chance is marked with the dashed line at 50% classification accuracy. Significantly above-chance accuracy values are marked by blue circles, while non-significant values are indicated by red crosses.